心理科学进展 2021, Vol. 29, No. 1, 112–122
Advances in Psychological Science

https://dx.doi.org/10.3724/SP.J.1042.2021.00112

# 词汇共现频率视角下语义联想效应及其神经机制*

李雨桐 隋 雪

(辽宁师范大学心理学院，大连 116029)

**摘 要** 词汇识别与记忆受到词汇之间语义联想关系的影响。鉴于自由联想方法建立词汇语义联想关系的局限性，研究者们提出用共现联想法建立词汇之间的语义联想关系。本文对词汇共现频率视角下的语义联想效应及其神经机制的研究进行分析发现，语义联想的丰富程度和关联程度影响词汇的加工；与少语义联想词相比，多语义联想词导致 P200 波幅增大和 N400 波幅减小；语义联想加工涉及的脑区主要是额上回和左内侧颞叶区域。未来研究需要探讨自由联想法和共现联想法建立语义联想关系的差异，基于共现联想法进一步地探究语义联想的行为规律和神经机制，并从共现联想的角度开展汉语语义联想效应及神经机制的研究。

**关键词** 共现频率，语义联想，自由联想，词汇识别

**分类号** B842

## 1 引言

在词汇学习与识别过程中，有些词汇，如桌子和椅子，医生和护士，教师和学生等，经常一起出现，当给被试呈现其中一个词时，被试会很容易地想到另外一个词，这种现象被称为语义联想(semantic associations) (Bentin et al., 1985)。研究者认为，有些词汇语义相关度较高，在心理词典的网络结构中距离较近，所以容易成为语义关联词(Baror & Bar, 2016; Otgaar et al., 2017; Zannino et al., 2018)。那么，如何确定词汇之间的语义联想关系？以往研究主要采用自由联想的方法(free association) (Lothane, 2018; Schachter, 2018)。该方法是将一个刺激词，如椅子，以视觉方式呈现给被试，要求被试尽快地说出他头脑中浮现的第一个词汇(Deese, 1959)，如果这个词汇是桌子，则“椅子−桌子”就是一对语义关联词。

研究者利用自由联想范式让被试对一个刺激词尽可能多地写出它的语义关联词，结果发现，词汇的语义关联词数量存在着明显差异(Geng et al., 2007; Shin et al., 2018)。有些词汇的语义关联词较多，如教师的语义关联词就有学生、学校、教室、教材、粉笔等；而有些词汇的语义关联词较少，如黄鼬，它的语义关联词仅有动物、鼬科等有限个词汇。研究还发现，一个词汇拥有语义关联词的数量影响个体对该词汇的学习。具体地说，拥有多个语义关联词的词汇容易学习(Lucas, 2000)。如果个体已经掌握的词汇中有很多词与待学习的新词有语义联想关系，那么语义联想会促进新词的学习。然而，当个体学习了一系列词汇(即一个词汇列表)，并继续学习新词时，如果新词与这个列表中多个词汇有语义联想关系，则对新词的错误记忆有所增加(Coane et al., 2016; Gallo, 2010)。

语义联想导致错误记忆的现象已被 DRM 范式(Deese-Roediger-McDermott) (Deese, 1959; Roediger & McDermott, 1995)实验多次证明(Jou et al., 2017; Jou et al., 2018; Wang et al., 2019)。这类实验多采用自由联想确认语义联想词对，再用 DRM 范式探讨语义联想对词汇学习的影响。实验分为学习和测验两个阶段。它的实验材料是 DRM 词表。经典的 DRM 词表由有语义联想关系的词组成，共计 36 个词表，每个词表 16 个词，其中一个词为关键诱饵(critical lure)。在学习阶段，每个词表依

收稿日期：2020-06-09
* 辽宁省教育厅自然科学基金青年项目(L201783637)，教育部人文社会科学规划基金项目(19YJA190005)，国家自然科学基金面上项目(31971036)资助。本研究是博士后部分研究工作(243514)。
通信作者：隋雪，E-mail: suixue88@163.com

次呈现关键诱饵以外的 15 个词，随后在测验阶段，呈现部分学习过的项目、无关项目(即学习阶段未呈现的项目)和关键诱饵，让被试判断这些项目在学习阶段是否呈现过。一系列实验发现，被试对关键诱饵的错误识别率较高(Otgaar et al., 2017; Woods & Dewhurst, 2019; Zhang et al., 2017)，说明语义联想影响个体对词汇的学习和再认。

然而，有研究者认为，采用自由联想方法确认语义联想词对有一定的局限性。首先，让大学生对一个词汇进行自由联想来确定语义关联词，尽管词对的语义接近且相关度较高，但是它与现实情境中语言的使用频率不一定一致。也就是说，该方法确定的联想程度高的词汇不一定在语境中经常关联着出现。其次，另有研究者质疑了自由联想方法建立词汇语义联想关系的正确性(Hofmann & Jacobs, 2014)。第一，自由联想是个体依赖于内部认知过程的一种行为表现，它本身是因变量。把这种个体行为表现(因变量)作为自变量去预测其他行为(如学习效果)，属于循环论证。第二，自由联想依赖于被试外显的自我报告，报告的结果受被试样本影响，主观性较强，且在报告过程中被试还可能遗漏一定数量或一定类型的联想词汇。

针对这些问题，研究者们提出用语境中词汇共现频率[1]确定词汇之间的语义联想关系，这种确定词汇之间语义联想关系的方法，被称为共现联想法(Hofmann et al., 2018)。共现(co-occurrence)是指在语言使用过程中两个词汇同时出现在同一个语境的现象。例如，在句子“中国是一个负责任、有担当的大国”中，中国与责任、中国与担当等就是共现关系。这种共现是在语境中共同出现，一个语境都是表达一种思想、观点，具有高度的严谨性，共现的词汇之间存在着一定关联。Hebb 的学习理论认为，两个词汇频繁地出现在同一语境，可以认定这两个词汇具有语义联想关系(Hebb, 1949; Rapp, 2002)。一旦两个词汇在语境中的共现频率具有显著的统计学意义，则这两个词汇就是语义关联词(Dunning, 1993)。共现联想法确定语义关联词更客观，语义关联词数量与现实情境中语言的使用频率也更为一致。这也说明共现联想法确定词汇之间联想关系更客观，这种方法确定的联想词汇加工的差异更能反映联想的作用。

鉴于共现联想法是另一种确定词汇之间语义联想关系的方法，本文旨在从共现频率视角对语义联想研究的成果进行综述，然后对未来研究进行展望，包括：自由联想法和共现联想法的区别与联系，共现联想法适用于汉语语义联想研究的可能性，并提出了从共现频率视角探究语义联想对汉语词汇学习影响的研究构想。

## 2　共现频率视角下词汇语义联想效应研究

### 2.1　词汇语义联想关系的确定

与自由联想方法不同，共现频率确定语义联想关系是通过对具体语境中语言使用频率的统计来确定的。例如在特定范围语境(M)中，词汇 A 出现频率为 $p(A)$，词汇 B 出现频率为 $p(B)$，从理论上讲，词汇 A 和词汇 B 的共现频率应该是 $P(A) \times P(B)$，如果实际上词汇 A 和词汇 B 的共现频率显著大于 $P(A) \times P(B)$，则确定词汇 A、B 为语义关联词(Dunning, 1993)。两个词汇在语境中共现频率越高，语义关联程度越高。

词汇的共现包括一阶共现和二阶共现两种类型(Hofmann & Jacobs, 2014)。一阶共现(first-order occurrence)是指两个词汇频繁地出现在同一个语境中，其共现频率显著大于两个词汇共现的理论概率。若两个词汇是一阶共现，则这两个词汇具有直接语义联想关系(Dunning, 1993)，如图 1 所示，“医生−医院”、“医生−处方”、“医生−护士”等就是一阶共现，具有直接语义联想关系。二阶共现(second-order occurrence)是指两个词汇频繁地出现在相似语境中，它们有相同的一阶共现词汇。二阶共现词汇具有间接语义联想关系(Rapp, 2002)，如图 1 中的“医生−护士”经常出现在相似语境中，有相同的一阶共现词汇(医院、手术、病

[1] 共现概率是针对两个词汇的，是指在某一具体语境中同时出现的概率。转换概率(transitional probability)是针对两个相邻词汇，一个词与另一个词共同出现的概率。转换概率可以分为，一个词与前一个词共同出现的概率和一个词与后一个词共同出现的概率。词汇的共现频率统计时包含了两个相邻词汇的转换概率，所以，两者是包含关系。互信息(mutual information)是基于共现频率来度量一个随机变量减少另外一个随机变量不确定程度的指标。两个词汇的互信息越大，相互之间减少不确定性程度越大。语境多样性(contextual diversity)是针对某一个词汇，通过在某一具体语境中，某一个词汇所出现的篇章数来测定。共现频率反映了两个词汇共同出现的概率问题，而语境多样性反映了单个词汇出现篇章的数量问题，二者针对的对象不同。

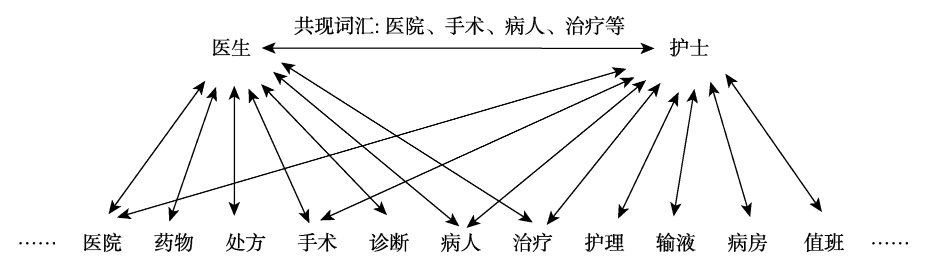


图 1 词汇之间的语义联想关系(数据来源于莱比锡语料库(见 2.3 节))

人、治疗等)。词汇之间可以具有一阶共现、二阶共现，或者兼而有之，如“医生–护士”既是一阶共现词汇又是二阶共现词汇，它们既有直接语义联想关系又有间接语义联想关系。

研究者也从计算语言学角度界定了词汇的一阶共现和二阶共现。一阶共现和二阶共现分别描述了词汇之间的组合关系(syntagmatic relation)和聚合关系(paradigmatic relation)。组合关系强调的是具有语义联想关系的词汇在语境中经常同时出现。而聚合关系强调的是具有语义联想关系的词汇经常在相似的语境中出现，在同一语境中可以互相替换(Biemann & Riedl, 2013; Namaziandost et al., 2018)。

总之，这种基于语境中词汇共现来建立词汇语义联想关系的方法，不再依赖于人类的行为表现，比基于主体自由联想建立的语义联想更全面，也避免了自由联想过于主观的不足。该方法为语义联想研究提供了新的视角，能够与自由联想方法相结合，从多个角度探讨语义联想效应的行为规律和神经机制。

## 2.2 词汇语义联想效应的理论解释

为了解释语义联想对词汇识别与再认的影响，研究者提出了联想读出模型(Associative Read-Out Model, AROM) (Hofmann & Jacobs, 2014; Hofmann et al., 2011)。AROM 是第一个增加了语义层的交互激活模型(interactive activation model, IAM) (McClelland & Rumelhart, 1981), AROM 是由 4 个水平组成，较低的三个水平对应交互激活模型(IAM)中特征水平、字母水平和单词水平。AROM 认为，视觉词汇识别是一个双向的过程，涉及到从词汇特征获得自下而上的信息和从对语义、词汇、字母成分的自上而下的处理。在特征水平上，当某个字母的某个特征被检测到时，激活的兴奋可以传送给包含这一特征的字母单元，同时将抑制传送给不包含这一特征的字母单元。在字母水平上，当处于词中某一特殊位置的字母被识别时，抑制将传送到其他字母，同时将兴奋传送到单词水平上所有在这个位置上包含该字母的单词。在单词水平上，被激活的单词一方面提高了这个单词组成字母的激活水平并抑制其他字母单元的激活，另一方面又将激活的兴奋传送给语义水平中与该单词单元对应的语义单元。在 AROM 语义水平上，一旦目标语义单元被激活，它就会顺次激活与之相关的所有语义单元，并抑制其他无关联语义单元的激活。目标语义单元激活相关语义单元的路径就是基于共现频率的语义联想关系，包括直接语义联想关系和间接语义联想关系，它们是人类联想记忆系统中的稳定结构(Hebb, 1949)。与此同时，这些被激活的语义单元也会顺次激活目标语义单元，被激活的语义单元数目越多，它们对目标语义单元的激活水平越高(Siew & Vitevitch, 2016)。因此，每个水平上特定单元的激活是导致竞争单元抑制的原因。AROM 从语义单元之间联想激活的扩散机制角度，能够解释语义联想对词汇识别与学习的加工过程。

另外，AROM 通过增加语义层，扩展了 IAM 对词汇识别加工过程的解释。IAM 将词汇识别分为由低到高的三个水平，分别是特征水平、字母水平和单词水平。IAM 中词汇识别的加工过程是通过激活和抑制的扩散在不同层级之间的传递来实现的。基于 IAM，后续研究者又提出了很多模型，这一类模型被统称为 IAMs，如 DROM (Dual Read-Out Model) (Grainger & Jacobs, 1994), MROM (Multiple Read-Out Model) (Grainger & Jacobs, 1996), DRC (Dual-Route Cascaded Model) (Coltheart et al., 2001), CDP++ (Connectionist Dual-Process Model) (Perry et al., 2007, 2010)等。然而，IAMs 忽视了词汇本身的语义作用，以及每个词汇还有语义关联词，这些语义关联词也会对词汇的识别与学习产生影响。因此，AROM 在 IAM 基础上，通

过增加语义层，考虑了词汇本身的语义作用以及词汇之间的联想关系，能更好地解释词汇识别与学习的加工现象。

### 2.3　词汇共现频率计算依赖的语料库

共现频率计算依赖于词汇出现的语境，在不同的语言环境中，语言习惯不同，词语搭配不同，共现频率也随之变化。莱比锡语料库(Leipzig Corpora Collection)提供了基于共现频率的语义联想关系的词汇信息(Goldhahn, 2012)。该语料库可以免费在线访问(http://corpora.uni-Leipzig.de)，它包含 250 多种语言的词典及其词汇的统计信息。每种语言抓取某一时间内互联网上的新闻、网页、维基百科等资源，并对资源中词汇进行统计。统计信息包括词频，词汇所在语句，具有统计学意义的共现词汇信息，词汇语义联想关系的网络可视化数据等。

汉语词汇的语义联想关系信息，是基于 2007~2009 年网上发布的新闻为资源进行统计的。该资源包括约 1900 多万个语句, 5.7 亿个词汇。莱比锡语料库还提供了数据下载功能，可以获得基于一定数量语句对词汇语义联想关系的统计信息。如，网站中可以下载的包含句子最多的数据集，共有 10 万个语句，约 24 万个词汇，提供了各个词汇在 10 万个语句中的词频、词汇之间共现次数、词汇之间基于对数似然比的显著性统计信息等，便于研究者编制实验材料。

### 2.4　基于共现频率的语义联想效应的行为学研究

研究者利用启动范式及真假词判断任务对语义联想的作用进行了研究(Roelke et al., 2018)。他们从莱比锡语料库中筛选出 300 个词，并编制了 100 个非词，组成了“启动词−目标词”词对作为实验材料，其中“词−词”和“词−非词”各 100 对。“词−词”中启动词和目标词的关系有 4 种，直接语义联想关系，间接语义联想关系，同时具有直接和间接的语义联想关系和无联想关系。该研究中语义联想关系的确定方法如下：直接语义联想关系由共现频率的对数来计算，间接语义联想关系由两个词汇共有的一阶共现词汇的数量来计算。结果发现，启动词与目标词同时存在直接和间接语义联想关系，与无联想关系比，判断的反应时更短，正确率更高。这证明了词汇判断中存在着语义联想效应。

有研究者(Hofmann et al., 2011)利用 DRM 范式探讨了语义联想的丰富程度对词汇识别与再认的影响。从莱比锡语料库挑选出 160 个德语名词作为实验材料。计算每个词与其他 159 个词汇语义联想关系的数量，如果大于 8 个，则称该词为多语义联想词；若少于 8 个，则称该词为少语义联想词。这 160 个词中，多语义联想词和少语义联想词各 80 个，学习阶段、测验阶段每种词各 40 个。学习阶段出现的词汇为旧词(80 个)，测验阶段新出现的词汇为新词(80 个)。研究发现，无论是新词还是旧词，被试对多语义联想词做出旧词反应的次数要显著高于少语义联想词。研究者(2011)认为，特定词汇(椅子)的呈现，会激活 AROM 语义层中它的语义关联词(如桌子、椅垫等)，这些语义关联词会导致该词的进一步激活，并且语义关联词越多，该词被激活的水平越高。由于多语义联想词的激活水平高，被试对其判定为旧词的可能性增大，做出旧词反应的次数也随之增多。因此，多语义联想词比少语义联想词的旧词反应更多。对于旧词，多语义联想词比少语义联想词的错误识别率低，而对于新词，则多语义联想词比少语义联想词的错误识别率高。

有研究还发现，多语义联想旧词比少语义联想旧词的反应时短，多语义联想新词比少语义联想新词的反应时长(Stuellein et al., 2016)。Braun 等人(2006)认为，呈现的刺激词是通过激活记忆中部分正字法相邻词汇，从而产生了整体的词汇激活(global lexical activity, GLA), GLA 越高，个体越能快速地做出正确的行为反应。与有少量正字法相邻词汇的词相比，有大量正字法相邻词汇的词的 GLA 更高，其识别的反应时更短(Hawelka et al., 2013)。因此，对于旧词，多语义联想词比少语义联想词的 GLA 更高，则多语义联想旧词比少语义联想旧词做出旧词的反应时更短。利用 GLA 还可以解释多语义联想新词比少语义联想新词反应时更长的现象。有研究发现，有大量正字法相邻词汇的非词比有少量正字法相邻词汇的非词的加工时间长(Braun et al., 2006)。其原因在于，早期视觉词汇加工过程中，非词要激活有正字法相似词汇表征的假设心理词汇。一个非词的正字法相邻词汇越多，它激活的假设心理词汇越多，这个非词的 GLA 可看作是它所激活的假设心理词汇的激活水平之和，整体词汇激活所需时间可通过每个假设心理词汇激活过程所需时间之和来计算

(Braun et al., 2015; Braun et al., 2019)。由于新词在学习阶段未被呈现过，所以个体需要激活该词的语义关联词来对新词进行学习和再认。那么，激活的语义关联词数目越多，新词的 GLA 越高，整体词汇激活过程所需时间越长，从而多语义联想新词比少语义联想新词的反应时间更长。

在一项使用真假词判断任务对 7 岁和 9 岁儿童的研究中，还发现了语义联想的关联程度的作用。研究者选取 320 个词作为实验材料，其中从莱比锡语料库筛选 240 个名词，这 240 个词中 160 个词作为启动词，80 个词作为目标词，另外 80 个目标词由 40 个可发音非词和 40 个随机字母序列非词组成。启动词和目标词组成词对的语义联想关系分为：强直接语义联想关系，弱直接语义联想关系，强间接语义联想关系和弱间接语义联想关系(Franke et al., 2017)。结果发现，两组儿童对强直接语义联想关系词对比弱直接语义联想关系词对识别的错误率低。而对间接语义联想关系，7 岁儿童对强关联词对比弱关联词对识别的错误率高，9 岁儿童则对强关联词对比弱关联词对识别的错误率低。这表明，儿童存在着从直接语义联想关系词对到间接语义联想关系词对识别的发展趋势，9 岁组儿童已经表现出直接语义联想效应和间接语义联想效应同时存在。

总之，语义联想的丰富程度影响词汇识别与再认。学习过的词汇，多语义联想词比少语义联想词识别的时间短，准确率高。而再认判断时出现的新词，多语义联想词比少语义联想词更容易被当做见过，对应地识别时间长，正确率低。因此，多语义联想增加新词的错误记忆，而对减少旧词错误记忆具有促进作用。另外，无论是直接语义联想关系还是间接语义联想关系，均表现出强关联词对比弱关联词对的识别错误率低。可见，语义联想的关联程度也会影响词汇加工过程。

**2.5 语义联想效应的脑电及脑成像研究**

通过共现频率确定词汇的语义联想关系，然后利用 DRM 范式研究发现了语义联想影响词汇的学习和再认，其神经机制也得到了研究者们的关注和进一步地探究。事件相关电位(event-related potential, ERP) (Li et al., 2017; Park & Donaldson, 2016)的研究表明，语义联想的丰富程度不同，导致 P200 和 N400 的波幅改变(Stuellein et al., 2016)。Stuellein 等人(2016)从莱比锡语料库挑选出 160 个德语名词，根据每个词与其他 159 个词的语义联想数量是否大于 8，将这 160 个词分为 4 类，分别是多语义联想旧词、少语义联想旧词、多语义联想新词和少语义联想新词。首先，要求被试学习 80 个旧词(多语义联想词和少语义联想词各 40 个)，然后，在测验阶段判断呈现的 160 个词汇是新词还是旧词，同时记录脑电。结果发现，旧词比新词产生更小的 N400 波幅和更大的 P200 波幅，多语义联想词比少语义联想词产生更小的 N400 波幅和更大的 P200 波幅。

研究者认为，从语义水平到正字法水平的自上而下的加工过程可以用 ERP 成分来说明(Jared et al., 2017; Midgley et al., 2009)，其中 P200 成分能反映词频效应中通达假设心理词汇时间的上限(Dambacher et al., 2006; Qu et al., 2016)。Stuellein 等人(2016)认为，词汇通达开始于 150~215 ms 之间。当刺激词激活的语义单元数目越多，这些语义单元对刺激词的激活水平越高，P200 波幅越大。同理，旧词比新词的激活水平高，多语义联想词比少语义联想词的激活水平高，旧词和多语义联想词加快了词汇识别与再认中词汇通达的过程，伴随着 P200 波幅增大。可见，P200 验证了从语义水平到单词水平的自上而下的词汇加工过程(Hofmann & Jacobs, 2014)。有关 N400 波幅的改变，研究者认为其原因在于，拥有较多语义关联词的词汇，其本身的激活状态较高，熟悉度也较高，加工相对容易，所以 N400 波幅较小。这说明，多语义联想以及高熟悉度导致词汇加工难度下降。由此可见，N400 与词汇的语义激活和熟悉度评价有关(Strozak et al., 2016; Voss & Federmeier, 2011)。总之，语义联想通过词汇激活加速了词汇通达，提高了词汇的熟悉程度，进而影响词汇的识别与再认。

研究者也对语义联想对词汇学习的研究进行了脑成像研究。Kuchinke 等人(2013)从莱比锡语料库挑选出 256 个德语名词，根据每个词与其他词的语义联想数量是否大于 10，将这 256 个词汇分为 4 类：多语义联想旧词、少语义联想旧词、多语义联想新词和少语义联想新词。利用 DRM 范式进行神经成像研究，结果发现，语义联想有利于记忆，多语义联想词比少语义联想词对左半球额上回(superior frontal gyrus, SFG)有更强的激活；多语义联想词引起左侧海马体更大程度的激活，

当被试能够高度自信地确认少语义联想词为新词和旧词时，这两类词汇也能导致海马体激活。有研究者认为，左侧海马体与情景记忆中回忆(Diana et al., 2007; Eichenbaum et al., 2007; Ross et al., 2017)和记忆强度(Smith et al., 2011; Wixted & Squire, 2010)有关。多语义联想词能够提高词汇的熟悉度，促进记忆，在回忆时快速做出反应，引起了左侧海马体更大程度的激活。而对于少语义联想词，少语义联想旧词做出旧词反应时需要在回忆过程中检索数量有限的语义联想词汇，大量的检索活动导致海马体更强的激活。少语义联想新词在加工过程中需要通过编码与其他词汇建立语义联想关系，这一过程使得海马体自动切换到编码模式，增强了后海马体和海马旁回的共同激活(Wixted et al., 2018)。研究还发现，这种语义联想程度和新旧词的交互作用，也导致小脑激活程度的改变(Kuchinke et al., 2013)。有研究者认为，小脑在情景记忆、语义记忆和语言处理中具有重要作用(Pleger & Timmann, 2018; Stoodley & Schmahmann, 2009)。

总之，语义联想相关的脑区是左内侧颞叶(left medial temporal lobe, left-MTL)区域，该区域包括前后海马体(anterior and posterior hippocampus)和邻近的海马旁回(adjacent parahippocampus)，以及左半球额上回和小脑区域。多语义联想词导致左半球额上回和左侧海马体更大程度的激活。同样地，少语义联想词在回忆过程需要大量的词汇检索以及新词的加工与编码，导致后海马体和海马旁回更强的激活。

## 3　小结与展望

现有研究主要采用自由联想法和共现联想法来确认语义联想词对，再用 DRM 范式探讨语义联想对词汇学习的影响。已有研究证明，共现联想法可行有效(Hofmann et al., 2018)，它可用于语义联想对词汇识别与学习的影响研究，还能生成满足更高需求的实验材料，使得语义联想研究可以探讨更为复杂的问题(Mandera et al., 2017; Mikolov et al., 2013)。如，通过共现联想法编制实验材料进行启动研究，直接语义联想关系的启动效应主要来自于联想启动，间接语义联想关系的启动效应主要来自于语义启动，探讨语义启动和联想启动是否可以分离。然而，共现联想法确定词汇之间的语义联想关系，从词汇共现频率视角对语义联想影响词汇识别与学习的研究还需要更多的实验进行验证。

第一，共现联想法和自由联想法建立语义联想关系的差异。自由联想确定的词对，两个词汇语义相关度较高，在心理词典的网络结构中距离较近。但是，自由联想过于依赖主体的知识与经验，主体自由联想时不可避免地会遗漏一定数量或一定类型的语义关联词。而共现联想法，是在语言使用过程中根据两个词汇共现同一语境来建立词汇之间的语义联想关系，它反映了人们语言使用的习惯和特点。该方法比主体自由联想确定的语义关联词更全面，也避免了自由联想过于主观的不足。当然，该方法也会将一些相关度不高、心理词典的网络结构中距离较远的词汇捆绑在一起，如图 1 中“护士–值班”。那么，哪种方法更适用于语义联想研究，或者说，哪种方法更适用于某一问题情境下的语义联想研究，是值得未来研究关注的问题。

第二，基于共现频率进行语义联想效应的研究，存在结论的不确定性。在行为学数据上，已有研究表明，多语义联想新词比少语义联想新词反应时长，这与早期研究结果不一致(Wagenmakers et al., 2008)。同样地，在神经机制上，词汇更大的语义激活导致 N400 波幅减少(Stuellein et al., 2016)，这也与之前研究结论不太符合。有研究在词汇判定任务中发现，词汇具有多个正字法相邻词和语义关联相邻词均会引起 N400 波幅增大(Müller et al., 2010)，而且有多个正字法相邻词的非词也会导致 N400 波幅增大(Meade et al., 2018)，并且词汇的激活水平与 N400 波幅呈正相关(Braun et al., 2006; Meade et al., 2019)。有研究者认为，可能是不同的任务(词汇判定任务和词汇再认任务)需求导致上述结论相悖。词汇识别和语义处理是两个不同的加工过程，这两个过程可能同时引发 N400 波幅的变化(Molinaro et al., 2010)。词汇判定任务中语义竞争过程可能导致 N400 波幅增大，而词汇再认任务中词汇的语义联想和熟悉度可能导致 N400 波幅减小。未来研究还需要进一步明确基于共现频率语义联想效应的行为规律和神经机制。此外，语义联想涉及复杂的心理过程，它是由多个脑区共同作用的结果。语义激活可能导致后颞中回(Hofmann et al., 2009)、额下回 IFG

(Hofmann & Jacobs, 2014)、左内侧颞叶区域和小脑(Stoodley & Schmahmann, 2009)等激活，但究竟哪些脑区与语义联想的丰富程度和关联程度有关，哪些脑区的共同作用影响语义联想对词汇识别与再认，还需要未来研究进一步地探讨。

第三，语义联想效应解释模型的普适性。语义联想效应的解释模型主要依赖于 AROM，它是第一个加上语义层的 IAM。IAMs 和 AROM 都是计算模型，这些模型是基于事先设定的规则，来解释词汇识别和学习的加工过程。当实验结果与模型解释一致时，利用模型的规则可以说明词汇识别和学习的加工机制。但是，现有研究结果存在着反应时和 N400 效应不一致的结论，该模型不能给出很好的解释，说明模型不能解释所有的词汇加工现象，其普适性有待提高。Hofmann 也提到 AROM 语义水平中词汇之间语义联想关系存在着语义特征关系的重叠，建议将上述两种关系添加到 AROM 中，从而提高模型的解释力(Stuellein et al., 2016)。

第四，词汇共现联想法应用于汉语语义联想效应研究的可能性。目前，共现联想法主要应用于英语和德语等语义联想效应研究。英语和德语等为印欧语系，属于低语境文化语言。如 WEDDING 的语义联想词，主体首先想到的是同义词 MARRIAGE，然后才是与之相关的 WIDOW 和 THRONE (Hofmann & Jacobs, 2014)。而汉语为汉藏语系，属于高语境文化语言。对于词汇“护士”，主体很容易想到语境中与之频繁共现的“医生”、“医院”等词汇。因此，汉语的语义联想更依赖于语境，共现联想法又是基于语境提出的确定词汇之间语义联想关系的方法，该方法对于汉语语义联想研究可能更为敏感。

那么，汉语语义联想是否对词汇识别与再认产生影响？有研究者认为，基于词汇共现频率来确定的语义关联词对，由于两个词汇在语言使用中频繁共现，从而导致了二者有较强的词汇联结(Trevor, 2008)，词汇的识别与再认和记忆特定的词汇联结密切相关，受到语言使用中特定词汇搭配经验的影响，这一加工过程并无自上而下的语义作用对其产生影响。另有研究发现，若两个词仅在语义上关联，却未在实际的语言运用中呈高频共现，那么，词汇识别并不能得到有效地促进(Ellis et al., 2009)。也就是说，语义联想对词汇识别与再认是否产生影响，或者与其他变量共同作用才能影响词汇识别与再认，目前尚无定论。此外，基于共现联想法对拼音文字得出的语义联想效应结论是否具有跨语言一致性，汉语语义联想效应是否具有差异性规律等，也是有待探究的问题。为此，针对共现联想法应用于汉语语义联想效应研究，我们有如下构想：

构想一，基于共现联想法检验词汇语义联想对汉语词汇识别与再认的影响，并探讨其神经机制。根据拼音文字的现有研究成果，以名词作为研究对象，并针对汉语名词的分类，分别探讨单字词和双字词的语义联想对词汇识别与再认的影响和神经机制。(1)针对单字词，首先，采用共现频率方法筛选单字词实验材料，并在词汇再认记忆任务中验证是否具有语义联想效应。若有，说明汉语单字词存在语义联想效应，共现频率方法适用于汉语的语义联想研究；若无，则采用自由联想方法筛选实验材料进行重新实验。如果实验结果出现语义联想效应，那么，对于汉语单字词，自由联想方法更适用于语义联想效应研究；反之，两种方法一致得出汉语单字词的语义联想对词汇再认无影响的结论。(2)与单字词相比，双字词的语义更明确，与语义关联词的关联程度更高。我们假定，双字词有语义联想效应，并且伴随着 P200 和 N400 的波幅变化。因此，需要进一步探讨汉语与其他语种产生的语义联想效应是否相同，背后的神经机制是否一致。

构想二，探讨相关因素在语义联想影响词汇识别和再认中的作用及机制。相关因素有记忆负荷、词汇的学习阶段和测验阶段的时间间隔，学习阶段的重复次数和学习方式等，探讨这些因素对语义联想效应的影响。特别需要注意的是，语义联想与其他变量的共同作用，如词频、具体性等，究竟是语义联想还是变量之间共同作用对词汇识别与再认产生影响，也是未来要关注的研究问题。

构想三，比较语义联想效应跨语言的共同性与差异性。基于共现联想法来探讨语义联想对词汇识别与再认的研究，主要应用于英语和德语等拼音文字，且已经取得了一定的研究成果，基于此，研究者也提出了语义联想效应的解释模型 AROM。汉语是表意文字，与拼音文字不同，语言差异是否也会在词汇联想效应中体现出来，联想是否影响汉语词汇识别与加工，进一步验证

AROM 模型合理性。

## 参考文献

Baror, S., & Bar, M. (2016). Associative activation and its relation to exploration and exploitation in the brain. *Psychological Science, 27*(6), 776–789.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343–355.

Biemann, C., & Riedl, M. (2013). Text: Now in 2D! A framework for lexical expansion with contextual similarity. *Journal of Language Modelling, 1*(1), 55–95.

Braun, M., Jacobs, A. M., Hahne, A., Ricker, B., Hofmann, M. J., & Hutzler, F. (2006). Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. *Brain Research, 1073-1074*, 431–439.

Braun, M., Jacobs, A. M., Richlan, F., Hawelka, S., Hutzler, F., & Kronbichler, M. (2015). Many neighbors are not silent. fMRI evidence for global lexical activity in visual word recognition. *Frontiers in Human Neuroscience, 9*, 423–423.

Braun, M., Kronbichler, M., Richlan, F., Hawelka, S., Hutzler, F., & Jacobs, A. M. (2019). A model-guided dissociation between subcortical and cortical contributions to word recognition. *Scientific Reports, 9*, 4506.

Coane, J. H., McBride, D. M., Termonen, M. -L., & Cutting, J. C. (2016). Categorical and associative relations increase false memory relative to purely associative relations. *Memory & cognition, 44*(1), 37–49.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.

Dambacher, M., Kliegl, R., Hofmann, M. J., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084*(1), 89–103.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences, 11*(9), 379–386.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19*(1), 61–74.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*(1), 123–152.

Ellis, N. C., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In U. Rõmer, & R. Schulze (Eds.), *Studies in Corpus Linguistics, Exploring the Lexis-Grammar Interface* (pp. 89–114). Amsterdam, Holand: John Benjamins.

Franke, N., Roelke, A., Radach, R. R., & Hofmann, M. J. (2017, July). *After braking comes hasting: reversed effects of indirect associations in 2nd and 4th graders.* Paper presented at the meeting of the Cognitive Science Society, London, England.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*(7), 833–848.

Geng, H., Qi, Y., Li, Y., Fan, S., Wu, Y., & Zhu, Y. (2007). Neurophysiological correlates of memory illusion in both encoding and retrieval phases. *Brain Research, 1136*, 154–168.

Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1158–1176.

Goldhahn, D., Eckart, T., Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection_ From 100 to 200 Languages. In N. Calzolari, K. Choukri, T. Declerck, M. Ugur Dogan, B. Maegaard, J. Mariani, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)* (pp. 3196–3200). European Language Resources Association (ELRA).

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*(3), 518–565.

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2013). Beyond single syllables: The effect of first syllable frequency and orthographic similarity on eye movements during silent reading. *Language and Cognitive Processes, 28*(8), 1134–1153.

Hebb, D. O. (Eds). (1949). *The organization of behavior: a neuropsychological theory.* New York: Wiley.

Hofmann, M. J., Biemann, C., Westbury, C., Murusidze, M., Conrad, M., & Jacobs, A. M. (2018). Simple co-occurrence statistics reproducibly predict association ratings. *Cognitive Science, 42*(7), 2287–2312.

Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience & Biobehavioral Reviews, 46*, 85–104.

Hofmann, M. J., Kuchinke, L., Biemann, C., Tamm, S., & Jacobs, A. M. (2011). Remembering words in context as

predicted by an associative read-out model. *Frontiers in Psychology, 2*, 11.

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L. H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience, 9*(4), 389–397.

Jared, D., Jouravlev, O., & Joanisse, M. F. (2017). The effect of semantic transparency on the processing of morphologically derived words: Evidence from decision latencies and event-related potentials. *Journal of Experimental Psychology: Learning, Memory and Cognition, 43*(3), 422–450.

Jou, J., Arredondo, M. L., Li, C., Escamilla, E. E., & Zuniga, R. (2017). The effects of increasing semantic-associate list length on the Deese-Roediger-McDermott false recognition memory: Dual false-memory process in retrieval from sub- and supraspan lists. *The Quarterly Journal of Experimental Psychology, 70*(10), 2076–2093.

Jou, J., Escamilla, E. E., Arredondo, M. L., Pena, L., Zuniga, R., Perez, M., & Garcia, C. (2018). The role of decision criterion in the Deese-Roediger-McDermott (DRM) false recognition memory: False memory falls and rises as a function of restriction on criterion setting. *The Quarterly Journal of Experimental Psychology, 71*(2), 499–521.

Kuchinke, L., Fritzemeier, S., Hofmann, M. J., & Jacobs, A. M. (2013). Neural correlates of episodic memory: Associative memory and confidence drive hippocampus activations. *Behavioural Brain Research, 254*, 92–101.

Li, B. B., Taylor, J. R., Wang, W., Gao, C. J., & Guo, C. Y. (2017). Electrophysiological signals associated with fluency of different levels of processing reveal multiple contributions to recognition memory. *Consciousness and Cognition, 53*, 1–13.

Lothane, H. Z. (2018). Free association as the foundation of the psychoanalytic method and psychoanalysis as a historical science. *Psychoanalytic Inquiry, 38*(6), 416–434.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review, 7*(4), 618–630.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language, 92*, 57–78.

Mcclelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, part I: An account of basic findings. *Psychological Review, 88*(88), 580–596.

Meade, G., Grainger, J., & Holcomb, P. J. (2019). Task modulates ERP effects of orthographic neighborhood for pseudowords but not words. *Neuropsychologia, 129*, 385–396.

Meade, G., Midgley, K. J., Dijkstra, T., & Holcomb, P. J. (2018). Cross-language neighborhood effects in learners indicative of an integrated lexicon. *Journal of Cognitive Neuroscience, 30*(1), 70–85.

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2009). Masked repetition and translation priming in second language learners: A window on the time-course of form and meaning activation using ERPs. *Psychophysiology, 46*(3), 551–565.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, May). *Efficient estimation of word representations in vector space*. Paper presented at the meeting of the international conference on learning representations, Scottsdale, USA.

Molinaro, N., Conrad, M., Barber, H. A., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience, 1*(1), 1–7.

Müller, O., Duñabeitia, J. A., & Carreiras, M. (2010). Orthographic and associative neighborhood density effects: What is shared, what is different? *Psychophysiology, 47*(3), 455–466.

Namaziandost, E., Shafiee, S., & Rasooyar, H. (2018). Paradigmatic relations and syntagmatic relations: Are they useful in learning grammatical structures? *Journal of Applied Linguistics and Language Research, 5*(5), 20–34.

Otgaar, H., Moldoveanu, G., Wang, J. Q., & Howe, M. L. (2017). Exploring the consequences of nonbelieved memories in the DRM paradigm. *Memory, 25*(7), 922–933.

Otgaar, H., Muris, P., Howe, M. L., & Merckelbach, H. (2017). What drives false memories in psychopathology? A case for associative activation. *Clinical Psychological Science, 5*(6), 1048–1069.

Park, J. L., & Donaldson, D. I. (2016). Investigating the relationship between implicit and explicit memory: Evidence that masked repetition priming speeds the onset of recollection. *NeuroImage, 139*, 8–16.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273–315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology, 61*(2), 106–151.

Pleger, B., & Timmann, D. (2018). The role of the human cerebellum in linguistic prediction, word generation and verbal working memory: Evidence from brain imaging, non-invasive cerebellar stimulation and lesion studies. *Neuropsychologia, 115*, 204–210.

Qu, Q. Q., Zhang, Q. F., & Damian, M. F. (2016). Tracking the time course of lexical access in orthographic production:

An event-related potential study of word frequency effects in written picture naming. *Brain and Language, 159*, 118–126.

Rapp, R. (2002, September). *The computation of word associations: comparing syntagmatic and paradigmatic approaches.* Paper presented at the meeting of the 19th international conference on Computational linguistics, Taipei, China.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & Hofmann, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychonomic Bulletin & Review, 25*(4), 1488–1493.

Ross, D. A., Sadil, P., Wilson, D. M., & Cowell, R. A. (2017). Hippocampal engagement during recall depends on memory content. *Cerebral Cortex, 28*(8), 2685–2698.

Schachter, J. (2018). Free association: From Freud to current use—the effects of training analysis on the use of free association. *Psychoanalytic Inquiry, 38*(6), 457–467.

Shin, J. -E., Suh, E. M., Eom, K., & Kim, H. S. (2018). What does “happiness” prompt in your mind? Culture, word choice, and experienced happiness. *Journal of Happiness Studies, 19*(3), 649–662.

Siew, C. S. Q., & Vitevitch, M. S. (2016). Spoken word recognition and serial recall of words from components in the phonological network. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42*(3), 394–410.

Smith, C. N., Wixted, J. T., & Squire, L. R. (2011). The hippocampus supports both recollection and familiarity when memories are strong. *The Journal of Neuroscience, 31*(44), 15693–15702.

Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *Neuroimage, 44*(2), 489–501.

Strozak, P., Abedzadeh, D., & Curran, T. (2016). Separating the FN400 and N400 potentials across recognition memory experiments. *Brain Research, 1635*, 41–60.

Stuellein, N., Radach, R. R., Jacobs, A. M., & Hofmann, M. J. (2016). No one way ticket from orthography to semantics in recognition memory: N400 and P200 effects of associations. *Brain Research, 1639*, 88–98.

Trevor, A. H. (Eds). (2008). *The psychology of language, from data to theory*. Hove and New York: Psychology Press.

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology, 48*(4), 532–546.

Wagenmakers, E. –J., Ratcliff, R., Gomez, P., & Mckoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*(1), 140–159.

Wang, J. Q., Otgaar, H., Howe, M. L., & Zhou, C. (2019). A self-reference false memory effect in the DRM paradigm: Evidence from eastern and western samples. *Memory & Cognition, 47*(1), 76–86.

Wixted, J. T., Goldinger, S. D., Squire, L. R., Kuhn, J. R., Papesh, M. H., Smith, K. A., … Steinmetz, P. N. (2018). Coding of episodic memory in the human hippocampus. *Proceedings of the National Academy of Sciences, 115*(5), 1093–1098.

Wixted, J. T., & Squire, L. R. (2010). The role of the human hippocampus in familiarity-based and recollection-based recognition memory. *Behavioural Brain Research, 215*, 197–208.

Woods, J. A., & Dewhurst, S. A. (2019). Putting false memories into context: The effects of odour contexts on correct and false recall. *Memory, 27*(3), 379–386.

Zannino, G. D., Perri, R., Teghil, A., Caltagirone, C., & Carlesimo, G. A. (2018). Associative agreement as a predictor of naming ability in alzheimer's disease: A case for the semantic nature of associative links. *Frontiers in Behavioral Neuroscience, 11*(261), 1–11.

Zhang, W. W., Gross, J., & Hayne, H. (2017). The effect of mood on false memory for emotional DRM word lists. *Cognition and Emotion, 31*(3), 526–537.

# Semantic association effect and its neural mechanism from the perspective of lexical co-occurrence frequency

LI Yutong, SUI Xue

(*Department of psychology, Liaoning Normal University, Dalian* 116029, *China*)

**Abstract:** Word recognition and memory are influenced by the semantic associative relationship between words. The free association method has some limitations in establishing lexical semantic associations. Researchers have proposed a new method called co-occurrence association. This study analyzed the semantic association effect and its neural mechanism from the perspective of lexical co-occurrence frequency. The results showed that the number and intensity of semantic association influenced the processing of words. Words with a large number of semantic associations led to an increase in the amplitude of P200 and a decrease in the amplitude of N400. The brain regions involved in semantic association processing are mainly the superior frontal gyrus and the left medial temporal lobe. Future research should discuss the differences between free and co-occurrence association in establishing a semantic association. Based on co-occurrence association, the behavioral characteristics and neural mechanisms of semantic association are further explored. Research on the semantic association effect and neural mechanism from the perspective of co-occurrence association is recommended.

**Key words:** co-occurrence frequency, semantic association, free association, word recognition